J M S K

Jurnal Matematika, Statistika, & Komputasi

http://journal.unhas.ac.id/index.php/jmsk

Vol. 4, No.1, 19-25, Juli 2007

ISBN:1858-1382

Suatu Tinjauan Tentang Generalized Estimating Equation

Muhammad Abdy*

Abstrak

Generalized Estimating Equation (GEE) yang pertama kali diperkenalkan oleh Liang dan Zeger pada tahun 1986 telah mendapat perhatian yang luas dari beberapa tahun terakhir dan beberapa perluasan dari metode tersebut telah dikembangkan. Pada tulisan ini akan disajikan metode GEE dan beberapa perluasannya, dan pada bagian terakhir dari tulisan ini akan diberikan contoh data yang akan dianalisis dengan menggunakan metode GEE.

Kata kunci: *generalized estimating equation, longitudinal, regresi*.

1. Pendahuluan

Tujuan utama dari analisis dalam sebagian besar model analisis regresi adalah mengetahui atau memeriksa pengaruh kovariat-kovariat tertentu terhadap variabel respon. Berbagai metode telah tersedia untuk mengestimasi parameter-parameter dalam model regresi dengan variabel respon yang berdistribusi normal dan kontinu. Akan tetapi, dalam aplikasi sering dijumpai variabel respon yang tidak normal dan kategorik sehingga akan memerlukan metode pendugaan yang lain untuk mengestimasi parameter-parameternya. *Generalized Linear Model* (GLM) yang diperkenalkan oleh *Nelder dan Wedderburn* pada tahun 1972 merupakan salah satu model regresi untuk menganalisis variabel respon yang diskrit maupun kontinu dan tidak perlu berdistribusi normal tetapi termasuk dalam distribusi keluarga eksponensial. Hubungan antara variabel respon dan kovariat dinyatakan dengan suatu fungsi yang disebut fungsi *link*. GLM mengasumsikan bahwa pengamatan adalah independen dan tidak mempertimbangkan adanya korelasi diantara outcome dari *n* buah pengamatan.

Pada pengumpulan data yang bersifat *longitudinal*, yaitu data *cross-sectional* yang diukur atau diamati secara berulang dalam urutan waktu, akan menyebabkan terjadinya interkorelasi diantara outcome-outcome. Metode pendugaan yang selama ini digunakan tidak akan valid lagi karena metode-metode tersebut mengasumsikan bahwa pengukuran outcome-outcome tersebut adalah independen. Untuk mengatasi adanya interkorelasi diantara outcome-outcome tersebut, maka GLM diperluas atau dikembangkan lagi sehingga muncul model-model marginal (*marginal models*) dan model-model pengaruh acak (*random effect models*).

Dalam model marginal, perhatian utama analisisnya adalah memodelkan ekspektasi marginal variabel respon. Korelasi diantara variabel outcome dimodelkan secara terpisah dan dipandang atau dianggap sebagai parameter penggangu (*nuisance parameters*). Model marginal pertama kali diperkenalkan oleh Zeger, Liang dan Self (1985), Liang dan Zeger (1986) dan Zeger dan Liang (1986). Dalam artikel-artikel tersebut, penulisnya memberikan definisi secara rinci tentang pendekatan tersebut dan memberikan beberapa contoh aplikasi. Pendekatan tersebut diistilahkan dengan *Generalized Estimating Equation* (GEE). Sampai sekarang ini GEE juga sudah mengalami perkembangan yang pesat dan sudah terjadi beberapa modifikasi untuk

* *Jurusan Matematika FMIPA Universitas Negeri Makassar*
Email: muh.abdy@unm.ac.id, abdy02@yahoo.com

mengakomodir berbagai macam permasalahan dalam data. Pemakaian metode GEE sudah sangat meluas diantaranya dalam bidang penelitian kesehatan dan genetika.

GEE 1

GEE merupakan perluasan dari GLM untuk mengakomodir data yang berkorelasi. Metode GEE memodelkan suatu fungsi yang diketahui dari ekspektasi marginal variabel respon sebagai fungsi linear dari satu atau lebih variabel prediktor. Dalam metode ini, kita tidak menspesifikasikan secara penuh distribusi bersama dari variabel-variabel respons, tetapi hanya menspesifikasikan fungsi *link*, hubungan antara mean dan varians respon serta struktur kovarians dari pengamatan dalam subjek. GEE orde satu merupakan GEE yang belum mengalami modifikasi.

Misalkan y_{it} adalah suatu prediktor respon dari n kelompok dengan T pengamatan untuk kelompok ke-i, i=1,2,…,n. Untuk setiap y_{it}, beberapa kovariat $\boldsymbol{x}_{it}$ tersedia, dimana elemen pertama dari $\boldsymbol{x}_{\mathrm{it}}$ adalah **1**, yang merupakan intersep. Data dapat dinyatakan dengan prediktor $\boldsymbol{y}_{\mathrm{i}}$ dan matriks $\boldsymbol{X}_i = (\boldsymbol{x}'_{i1}, \ \boldsymbol{x}'_{i2}, \ \ldots, \ \boldsymbol{x}'_{iT})$. Pasangan $(\boldsymbol{y}_i, \boldsymbol{X}_i)$ diasumsikan berdistirbusi identik independen. Misalkan $E(y_{it}|\ \boldsymbol{X}_i) = E(y_{it}|\boldsymbol{x}_{it}) = \boldsymbol{x}'_{it}\boldsymbol{\beta}$ **,** dimana $\boldsymbol{\beta}$ adalah suatu prediktor parameter berukuran (p×1) yang tidak diketahui. Selanjutnya diasumsikan bahwa matriks varians bersyarat dari $\boldsymbol{y}_i$ diberikan $\boldsymbol{X}_i$ adalah diketahui, yaitu kov$(\boldsymbol{y}_i|\boldsymbol{X}_i) = \boldsymbol{V}_i$, maka estimator model regresi linear multivariate dapat diperoleh melalui persamaan estimasi:

$$\boldsymbol{\mu}(\boldsymbol{\beta}) = \tfrac{1}{n}\boldsymbol{X}'\boldsymbol{V}^{-1}(\boldsymbol{y} - \mu) = \boldsymbol{0}\ , \qquad (1)$$

dimana $\boldsymbol{X}$ dan $\boldsymbol{y}$ masing-masing berasal dari matriks $\boldsymbol{X}_i$ dan prediktor $\boldsymbol{y}_i$, $\boldsymbol{V}$ matriks diagonal dari $\boldsymbol{V}_i$, $\boldsymbol{\mu}$ berasal dari prediktor $\boldsymbol{\mu}_i = \boldsymbol{\mu}_i(\boldsymbol{\beta})$. Estimator tersebut di atas tidak bias dan berdistribusi normal asimptotik. Matriks kovariansnya merupakan invers dari matrik informasi Fisher. Jika cov$(\boldsymbol{y}_i|\boldsymbol{X}_i) = \boldsymbol{\Omega}_i \neq \boldsymbol{V}_i$, maka estimator tersebut masih tak bias. Matriks variansnya berbentuk:

$$\boldsymbol{V}(\hat{\beta}) = (\boldsymbol{X}'\boldsymbol{V}^{-1}\boldsymbol{X})^{-1}(\boldsymbol{X}'\boldsymbol{V}^{-1}\boldsymbol{\Omega}\boldsymbol{V}^{-1}\boldsymbol{X})(\boldsymbol{X}'\boldsymbol{V}^{-1}\boldsymbol{X})^{-1} \qquad (2)$$

Suatu estimator dari (2) dapat diperoleh dengan mengganti $\boldsymbol{\Omega}_i$ dengan:

$$\boldsymbol{\Omega}_i = (\boldsymbol{y}_i - \hat{\mu}_{\boldsymbol{i}})(\boldsymbol{y}_i - \hat{\mu}_{\boldsymbol{i}})',$$

dimana $\hat{\mu}_{\boldsymbol{i}} = \boldsymbol{x}_i\hat{\boldsymbol{\beta}}$.

Matriks kovarian $\boldsymbol{V}_i$ yang fix dapat diganti dengan $V_i = V_i(\alpha)$ yang bergantung pada suatu parameter assosiasi α. Estimator $\hat{\boldsymbol{\beta}}$ diperoleh dengan prosedur dua tahap; yaitu, *tahap pertama* adalah mendapatkan parameter assosiasi $\hat{\alpha}$ untuk α dan *tahap kedua* adalah mendapatkan $\hat{\boldsymbol{\beta}}$ untuk β yang diestimasi. Pendekatan ini dapat digunakan untuk memodelkan ketergantungan di dalam kelompok pada model-model linear. Akan tetapi, model linear tidak dapat menangani semua persoalan-persoalan dalam aplikasi. Untuk pengamatan independen, GLM mempunyai fleksibilitas dalam memodelkan struktur mean dan varians. Dalam GLM tersebut, struktur mean adalah $E(y_{it}|\ \boldsymbol{X}_{it}) = \mu_{it} = g(\boldsymbol{x}'_{it}\beta)$, dimana g adalah suatu fungsi respons non-linear yang monoton. Invers dari fungsi g, yaitu g^{-1} dinamakan fungsi *link*. Suatu sifat penting dari GLM adalah relasi fungsional diantara mean dan varian, yaitu: $v_{it} = V(y_{it}|\ \boldsymbol{X}_{it}) = h(\mu_{it})$, dimana h disebut fungsi varians. Bentuk dari distribusi variabel respon akan menentukan bentuk dari fungsi *link* dan fungsi varians dari GLM.

Untuk pengamatan independen, prediktor parameter $\boldsymbol{\beta}$ diestimasi dengan menggunakan *metode kemungkinan maksimum*. Persamaan *kemungkinannya* berbentuk :

$$\boldsymbol{u}(\boldsymbol{\beta}) = \tfrac{1}{n}\sum_{i=1}^{n}\boldsymbol{D}'_i\boldsymbol{V}_i^{-1}(\boldsymbol{y}_i - \boldsymbol{\mu}_i) = \tfrac{1}{n}\boldsymbol{D}'\boldsymbol{V}^{-1}(\boldsymbol{y}-\boldsymbol{\mu}) = \boldsymbol{0} \qquad (3)$$

Dimana $\boldsymbol{D}_i = \dfrac{\partial \boldsymbol{\mu}_i}{\partial \boldsymbol{\beta}'}$ adalah matriks diagonal turunan pertama dan $\boldsymbol{V}_i = diag(v_{it})$ adalah matriks diagonal dari varians. Persamaan (3) biasa juga disebut dengan *Independence Estimating Equation* (IEE). Solusi dari (3) hanya ada untuk model linear dengan variabel respons berdistribusi normal. Jika variabel responnya tidak berdistribusi normal, maka harus diselesaikan secara iteratif. Estimator $\hat{\boldsymbol{\beta}}$ untuk (3) konsisten dan berdistribusi normal asymptotik dengan matriks kovarians kov($\hat{\boldsymbol{\beta}}$) = $(\boldsymbol{D}\boldsymbol{V}^{-1}\boldsymbol{D})^{-1}$ (Mc. Cullagh dan Nelder J.A. 1983)

Untuk pengamatan yang berkorelasi, matriks varians tidak dapat mempunyai suatu bentuk diagonal. Untuk data berkorelasi, Zeger dkk. (1985) mengusulkan penggunaan estimator varians yang kekar (robust):

$$\boldsymbol{V}(\hat{\beta}) = \left(\sum_{i=1}^{n} \hat{\boldsymbol{D}}_i \hat{\boldsymbol{V}}_i^{-1} \hat{\boldsymbol{D}}_i\right)^{-1} \left(\sum_{i=1}^{n} \hat{\boldsymbol{D}}_i' \hat{\boldsymbol{V}}_i^{-1} \Omega_i \hat{\boldsymbol{V}}_i^{-1} \hat{\boldsymbol{D}}_i\right) \left(\sum_{i=1}^{n} \hat{\boldsymbol{D}}_i' \hat{\boldsymbol{V}}_i \hat{\boldsymbol{D}}_i\right)^{-1} \tag{4}$$

Dalam persamaan tersebut di atas, $\hat{\Omega}$ merupakan matriks diagonal dari

$$\hat{\boldsymbol{\Omega}}_i = (\boldsymbol{y}_i - \hat{\boldsymbol{\mu}}_i)(\boldsymbol{y}_i - \hat{\boldsymbol{\mu}}_i)'$$

Estimasi tersebut di atas tidak akan efisien karena bentuk diagonal $\boldsymbol{V}_i$. Suatu pendekatan yang diusulkan oleh Liang dan Zeger (1986) dan Zeger dan Liang (1986) memberikan suatu estimasi yang lebih efisien dengan mengkombinasikan prosedur GLM dan FGLS. Prosedurnya adalah sebagai berikut: Pandang suatu model untuk ekspektasi dan varians dengan menggunakan pendekatan GLM. Dalam kasus ini, $\boldsymbol{V}_\mathrm{i}$ tidak perlu berbentuk matriks diagonal tetapi matriks kovarians sebaiknya mendekati matriks kovarians Ω_i yang sebenarnya. Untuk estimator $\hat{\boldsymbol{R}}_i$ dari matriks korelasi $\boldsymbol{y}_i$ diberikan $\boldsymbol{X}_i$, maka estimator untuk $\boldsymbol{V}_i$ berbentuk:

$$\hat{\boldsymbol{V}}_i = \hat{\boldsymbol{A}}_i^{1/2} \hat{\boldsymbol{R}}_i \hat{\boldsymbol{A}}_i^{1/2} \tag{5}$$

dimana $\hat{\boldsymbol{A}}_i^{1/2}$ adalah invers dari akar kuadrat matriks diagonal yang diestimasi dari varians V_{it}. $\hat{\boldsymbol{R}}_i$ merupakan matriks $T\times T$ definit positif yang menggambarkan struktur assosiasi. Matriks $\boldsymbol{R}_i$ biasa diistilahkan dengan “matriks korelasi kerja”. Matriks ini diestimasi dengan menggunakan metode momen. Jika $\hat{\boldsymbol{R}}_i$ merupakan matriks identitas, maka $\widehat{\boldsymbol{V}}_i$ pada persamaan (5) merupakan matriks diagonal yang akan sama dengan $\boldsymbol{V}_i$ dalam persamaan (3). Dengan matriks korelasi kerja $\hat{\boldsymbol{R}}_i$ dan matriks diagonal $\hat{\boldsymbol{A}}_i$, Generalized Estimating Equation (GEE) mempunyai bentuk:

$$\boldsymbol{\mu}(\boldsymbol{\beta}) = \tfrac{1}{\boldsymbol{n}} \sum_{\boldsymbol{i}=1}^{\boldsymbol{n}} \boldsymbol{D}'_{\boldsymbol{i}} \hat{\boldsymbol{V}}_{\boldsymbol{i}}^{-1} (\boldsymbol{y}_{\boldsymbol{i}} - \boldsymbol{\mu}_{\boldsymbol{i}}(\boldsymbol{\beta})) = \boldsymbol{0} \tag{6}$$

Jika $\boldsymbol{\beta}$ diestimasi menggunakan persamaan (6), maka $\hat{\boldsymbol{\beta}}$ akan konsisten jika $\mu_{it} = E(y_{it}| \boldsymbol{X}_{it}) = E(y_{it}| \boldsymbol{X}_i)$ dispesifikasikan dengan tepat. Demikian juga estimator $\hat{\boldsymbol{\beta}}$ akan berdistribusi normal asymptotik (Liang dan Zeger, 1986). Varians dapat diestimasi secara konsisten dengan estimator kekar (4) dan $\widehat{\boldsymbol{V}}_i$ diestimasi dengan menggunakan (5). Jika $\boldsymbol{V}_i$ dispesifikasikan dengan tepat, yaitu $\hat{\Omega}_\mathrm{i} = \boldsymbol{V}_i$, maka estimator $\hat{\boldsymbol{\beta}}$ akan efisien.

GEE 2

Dalam bagian ini akan dibahas persamaan estimasi yang memberikan suatu estimator konsisten dari mean. Persamaan estimasi tersebut disebut GEE 2. GEE 2 dipakai untuk estimasi secara bersamaan dari mean dan assosiasi (korelasi). GEE 2 mempunyai bentuk:

$$\mu\begin{pmatrix}\beta\\ \alpha\end{pmatrix}=\frac{1}{\mathrm{n}}\sum_{i=1}^{n}\begin{pmatrix}\frac{\partial\mu_i}{\partial\beta'} & 0\\ 0 & \frac{\partial\varphi_i}{\partial\alpha'}\end{pmatrix}'\begin{pmatrix}\boldsymbol{V}(\mathrm{y_i}) & 0\\ 0 & \boldsymbol{V}(\mathrm{z_i})\end{pmatrix}^{-1}\begin{pmatrix}\boldsymbol{y}_i-\mu_i\\ \boldsymbol{z}_i-\varphi_i\end{pmatrix}=0 \qquad (7)$$

Dari persamaan (7) tersebut terlihat bahwa matriks turunan pertama dan “matriks kovarians kerja” merupakan matriks yang berbentuk diagonal. Dengan demikian persamaan (7) dapat disederhanakan menjadi:

$$\mu\begin{pmatrix}\beta\\ \alpha\end{pmatrix}=\frac{1}{\mathrm{n}}\sum_{i=1}^{n}\begin{pmatrix}\frac{\partial\mu_i}{\partial\beta'} & \frac{\partial\mu_i}{\partial\alpha'}\\ \frac{\partial\varphi_i}{\partial\beta'} & \frac{\partial\varphi_i}{\partial\alpha'}\end{pmatrix}'\begin{pmatrix}\boldsymbol{V}(\boldsymbol{y}_i) & \mathrm{kov}(\boldsymbol{y}_i,\boldsymbol{z}_i)\\ \mathrm{kov}(\boldsymbol{z}_i,\boldsymbol{y}_i) & \boldsymbol{V}(\boldsymbol{z}_i)\end{pmatrix}^{-1}\begin{pmatrix}\boldsymbol{y}_i-\mu_i\\ \boldsymbol{z}_i-\varphi_i\end{pmatrix}=0 \qquad (8)$$

Dimana μ_i merupakan suatu fungsi dari parameter assosiasi α, jika $\frac{\partial\mu_i}{\partial\alpha'}\neq\mathbf{0}$. Akan tetapi asumsi ini tidak selalu dapat dilakukan. Lagi pula sangat sulit untuk menginterpretasi suatu vektor mean yang memasukkan α. Dalam kebanyakan aplikasi, nilai mean hanya didefinisikan sebagai suatu fungsi β, dan tidak bergantung pada α. Bentuk $\frac{\partial\mu_i}{\partial\alpha'}=\mathbf{0}$ dalam (8) mengakibatkan korelasi merupakan suatu fungsi dari β. Dalam aplikasi biasanya diasumsikan $\frac{\partial\varphi_i}{\partial\beta'}=\mathbf{0}$.

Sejauh ini kita hanya mendefinisikan persamaan estimasi dengan menggunakan korelasi sebagai ukuran assosiasi, padahal persamaan estimasi dapat juga didefinisikan dengan menggunakan matriks kovarians, yaitu:

$s_{itt'} = (y_{it}- \mu_{it})\,(y_{it'}- \mu_{it'})$ dan $\delta_{itt'} = E(s_{itt'}) = \mathrm{kov}(y_{it}\,,\, y_{it'})$

digunakan sebagai pengganti dari $z_{itt'}$ dan $\varphi_{itt'}$. Turunan pertama dan matriks varians kerja harus diubah yang sesuai. Sistem persamaan tersebut akan memenuhi yang berikut:

Jika $\boldsymbol{\mu}_i$ dan $\boldsymbol{\sigma}_i$ dispesifikasikan dengan tepat, sebagai fungsi dari α dan β, maka $\hat{\boldsymbol{\alpha}}$ dan $\hat{\boldsymbol{\beta}}$ dapat diestimasi secara konsisten dan estimator tersebut berdistribusi normal asimptotik. Sistem persamaan dengan matriks kovarians ini sangat jarang digunakan. Keuntungan menggunakan struktur kovarians untuk memodelkan korelasi adalah persamaan estimasi dapat diperoleh dengan suatu pendekatan *kemungkinan maksimum semu* (KMS). Dengan metode KMS ini, persamaan estimasi dapat juga dideskripsikan dengan menggunakan momen kedua. Hubungan diantara momen kedua dan log odd ratio dapat dideskripsikan dengan mudah (Birshop, Kienberg dan Holland, 1975). Jika persamaan (8) digunakan secara bersama-sama dengan log odd rasio dan mean serta struktur kovarian dispesifikasikan dengan benar, maka estimator konsisten $\hat{\boldsymbol{\alpha}}$ dan $\hat{\boldsymbol{\beta}}$ ada dan berdistribusi normal asimptotik. Apabila terdapat kekeliruan dalam menspesifikasikan α maka dapat menyebabkan suatu estimasi yang tidak konsisten dari β, karena β dan α diestimasi secara simultan.

Model orthogonal untuk parameter β dan α dan log odd ratio untuk struktur assosiasi dapat digunakan untuk mentransformasikan prosedur estimasi $\hat{\boldsymbol{\alpha}}$ dan $\hat{\boldsymbol{\beta}}$ secara simultan ke dalam suatu prosedur dua tahap. Estimator $\hat{\boldsymbol{\beta}}$ akan konsisten meskipun α tidak dispesifikasikan dengan benar. Pendekatan ini disebut *alternate logistic regression*, dimana logit-link digunakan sebagai fungsi link. Untuk lebih detail tentang alternate logistic regression, lihat Carey, Zeger dan Diggle, 1993.

Contoh Data

Berikut ini diberikan data dari suatu hasil penelitian tentang pengaruh polusi udara terhadap anak-anak (Diggle, P. J, dkk. 1995). Peneliti memeriksa 25 anak dan dicatat apakah anak-anak tesebut memperlihatkan gejala *wheezing*.

Tabel 1.

id	kota	8 thn		9 thn		10 thn		11 thn	
		merokok	gejala	merokok	gejala	merokok	gejala	merokok	gejala
1	A	0	1	0	1	0	1	0	0
2	A	2	1	2	1	2	1	1	0
3	A	2	1	2	0	1	0	0	0
4	B	0	0	1	1	1	1	0	0
5	A	0	0	1	0	1	0	1	0
6	B	0	1	0	0	0	0	0	1
7	A	1	1	1	1	0	1	0	0
8	B	1	0	1	0	1	0	2	0
9	B	2	1	2	0	1	1	1	0
10	A	0	0	0	0	0	0	1	0
11	A	1	1	0	0	0	0	0	1
12	B	0	0	0	0	0	0	0	0
13	A	2	1	2	1	1	0	0	1
14	B	0	1	0	1	0	0	0	0
15	A	2	0	0	0	0	0	2	1
16	B	1	0	1	0	0	0	1	0
17	B	0	0	0	1	0	1	1	1
18	A	1	1	2	1	0	0	1	0
19	A	2	1	1	0	0	1	0	0
20	B	0	0	0	1	0	1	0	0
21	A	1	0	1	0	1	0	2	1
22	B	0	1	0	1	0	0	0	0
23	B	1	1	1	0	0	1	0	0
24	A	1	0	1	1	1	1	2	1
25	B	0	1	0	0	0	0	0	0

Keterangan: **A** = kota yang mempunyai banyak pabrik baja
B = kota yang berbukit-bukit dan banyak pepohonan yang tumbuh

Pemeriksaan ini dilakukan empat kali, yaitu pada saat anak-anak berumur 8 tahun, 9 tahun, 10 tahun, dan 11 tahun. Responnya adalah gejala wheezing pada anak-anak yang bernilai 1 jika ada gejala dan 0 jika tidak ada gejala. Variabel prediktornya adalah kota tempat tinggal anak (dua kategori kota), tingkat keaktifan merokok dalam keluarga (bernilai 0, 1, dan 2), dan variabel umur anak. Datanya seperti dalam Tabel 1 dan sebagian hasil analisis seperti dalam Tabel 2. Data yang dihasilkan dalam penelitian ini merupakan data longitudinal, dan akan diolah dengan pendekatan metode GEE 1 dengan menggunakan SAS (proc. GENMOD). Dari Tabel 2, terlihat bahwa kota tempat tinggal bukanlah suatu faktor penyebab gejala wheezing pada anak-anak. Akan tetapi kebiasaan merokok dalam keluarga mempunyai pengaruh yang cukup signifikan, demikian juga umur anak mempengaruhi adanya gejala wheezing.

Tabel. 2

Analysis of GEE Parameter Estimates
Empirical Standard Error Estimates

Parameter	Estimate	Standard error	95% confidence Limits		Z	pr >|z|
Intercept	2.2615	2.0243	-1.7060	6.2290	1.12	0.26396
A	0.0418	0.5435	-1.0234	1.1070	0.80	0.9387
B	0.0000	0.0000	0.0000	0.0000	.	.
Umur	-0.3201	0.1884	-0.6894	0.0492	-1.70	0.0893
Merokok	0.6506	0.2821	0.0978	1.2035	2.31	0.0211

2. **Kesimpulan**

Pada bagian ini, akan diberikan beberapa rekomendasi untuk penggunaan metode GEE. Untuk penggunaan praktis, beberapa rekomendasi diperlukan untuk memutuskan apakah metode kemungkinan maksimum untuk distribusi multivariate ataukah metode GEE yang seharusnya digunakan. Secara umum, metode kemungkinan maksimum hanya digunakan jika distribusi lengkap y_i dengan syarat X_i dispesifikasikan dengan benar. GEE 1 menghasilkan estimator yang tidak konsisten untuk mean, jika tidak dispesifikasikan dengan benar. Penggunaan estimator kekar untuk varians harus digunakan jika ada kemungkinan terjadi kesalahan spesifikasi struktur assosiasi. Jika pemeriksaan assosiasi merupakan tujuan utama dari analisis, maka GEE 2 dapat digunakan karena estimasi dari struktur mean-nya konsisten meskipun assosiasi tidak dispesifikasikan dengan benar.

Daftar Pustaka

[1]. M. Abdy, 2003, ”*GEE: Suatu Metode Estimasi Parameter Model Regresi untuk Data Longitudinal”* – Makalah Seminar Nasional Statistika ITS – Pascasarjana UGM.

[2]. Y.M.M. Bishop, S.E. Fienberg dan P.W. Holland, 1975, “*Discrete Multivariate Analysis: Theory and Practice”*, MIT Press, Cambridge.

[3]. V. Cerey, S.L. Zeger and P.Diggle, 1983, *“Modeling Multivariate Binary Data with Alternating Logistic regression”*, *Biometrika* 80 : 517 – 526.

[4] P.J.Diggle, K.Y.Liang dan S.L. Zeger, 1995, “*Analysis of Longitudinal Data”*, Oxford University Press, Oxford.

[5] K.Y. Liang dan S.L. Zeger, 1986, “*Longitudinal Data Analysis Using GLM*”, *Biometrika* 73 : 13 – 22.

[6] P. Mc Cullagh dan J.A. Nelder,1983 *,"Quasi Likelihood Function, The Annals of Statistics",* 11 : 59 – 67.

[7] J. Nelder dan R.W.M. Wedderburn, 1972, "*Generalized Linear Models*", *Journal Royal Statistics Soceity A,* 135 : 370 384.

[8] S.L. Zeger dan K.Y. Liang, 1986, "Longitudinal *Data Analysis for Discrete and Continuous Outcomes*", *Biometrics* 42 : 121 – 130.